Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE INFLUENCE OF THE SEED UPON THE SIZE OF THE FRUIT IN STAPHYLEA. I

J. ARTHUR HARRIS

(WITH FOUR FIGURES)

A previous paper in this journal[1] reviewed the chief literature bearing on the theory that the size attained by a fruit is to some extent determined by a stimulus exerted by the developing seed. The purpose of the present contribution is to consider the data from two series comprising 3277 fruits of *Staphylea*, collected in a manner to permit of carrying the analysis somewhat farther than was possible with the *Cercis* material.

I. Nature of the characters considered

The fruit of *Staphylea trifolia* is familiar to botanists as a three-lobed, three-celled, membranaceous, inflated pod, producing from 0 to 4 bony seeds in each locule. Unfortunately it is hardly of the form one would select especially for an investigation of the size. Its considerable irregularity renders measurement difficult, and its lightness makes weighing tedious. Its use in the present study is in part a consequence of the ease with which the necessary data could be obtained in connection with other studies of fertility and fecundity in the species, and in part a result of special advantages which will be apparent to the reader later.

In a fruit whose shape and texture precludes any very accurate determination of size, the thing to be done is to take measurements on a scale coarse enough that minor irregularities will not be of much account. Having done this, we should not attach great importance to the exact values of the calculated constants, but look at them as merely rough approximations. Consistent results from a number of individual series will strengthen our confidence in the soundness of our conclusions, but if the relationships are of a very slight intensity, we should expect, for reasons well known to the statistician, different series of material to yield somewhat discordant results.

---

[1] HARRIS, J. ARTHUR, On the relationship between the length of the pod and fertility and fecundity in *Cercis*. BOT. GAZ. 50:117–127. 1910.

For a first study it seemed undesirable, considering the irregularity of the pod, to take measurements in fine units, and readings were made of the length of the fruit to the nearest 0.5 cm. The number of ovules formed and the number of seeds developing can be fairly easily counted.

II. Material

Shrubs of *Staphylea* growing in the "North American Tract" of the Missouri Botanical Garden furnished the material for this investigation. These are in part the individuals upon which a study of selective elimination[2] was based in the spring of 1908, and the numbers of the plants are comparable.

In 1906 the fruiting of *Staphylea* seemed to be quite normal, and a series of about 100 fruits each were obtained from 20 shrubs. In 1907 there was a severe frost which inhibited entirely the fruiting of some of the trees; altogether only 1218 fruits were obtained, as compared with 2059 in 1906. In 1906 the fruits were taken from shrubs 11–20, but in 1907 it was necessary to take shrubs 11–41 to get 16 individual series of material of suitable size. Only three-loculed fruits were used.

In the reduction of the data the biometric methods, now familiar in a general way to most working biologists, were used.[3] For the benefit of biologists who are inclined to be wary of mathematical symbols, I may explain that no difficulties of that kind will be met if a few simple ideas are kept in mind. For convenience and terseness of expression, the characters investigated are designated by letters: o = number of ovules per locule[4]; s = number of seeds per locule[4]; f = number of ovules failing to develop to mature seeds per locule[4]; p = position of fruit on the inflorescence; n = number of fruits per inflorescence; $r_{ls} > r_{lo}$ is to be read "the correlation (r being the symbol for correlation) between the length of the pod and the number of seeds developing is greater than the correlation between the length of the fruit and the number of ovules formed."

---

[2] HARRIS, J. ARTHUR, On the selective elimination occurring during the development of the fruit of *Staphylea*. Biometrika **7**:452–504. 1910.

[3] Unless especially noted, SHEPPARD's correction was not used for the second moment.

[4] Or occasionally total number per fruit, when this is indicated by the context.

III. Analysis of data

We first examine our data to ascertain whether there is a measurable degree of interdependence between the length of fruit and the number of ovules formed, and between the length of fruit and the number of seeds developing per locule.

A. DISCUSSION OF DATA FROM INDIVIDUAL SHRUBS

Perhaps the criticism, in these days, most generally directed at biometric work is that in the massing of large numbers of individuals into correlation tables, biological relationships which otherwise might be recognized are obscured. "The fundamental requisite for the validity of biometric constants is the homogeneity of the material upon which they are based," we are told. There is more than one word to be said on this point, but certainly it is always desirable to consider data in as minutely analyzed form as can be done without incurring too great dangers incident to the probable errors of random sampling.

From PEARSON'S investigations[5] we know that plant individuals of the same race are somewhat differentiated among themselves with respect to the characters of the organs which they produce. In short, they are *individual* really as well as nominally. The combination of series of fruits taken from a small number of plants might influence to some extent the correlation constants describing the relationship between the fertility of the fruit and its length. To free our results as far as possible from any such source of error, I first consider the relationship of the number of ovules formed (o) and the number of seeds developing (s) to the length of the fruit (l) in each of the 20 shrubs of the 1906 series, and each of the 16 of the 1907 collection. In the large series it is also of interest to compare the correlation for the length of fruit and number of ovules failing to develop into mature seeds (f). For these relationships 92 correlation tables are necessary. While experience has shown that it is desirable that all tables of data should be published, it really seems unreasonable to ask a publisher to print such a series. I retain my original tables, which are available to anyone desiring to consult them, and have deposited a duplicate set in the Library of the Missouri Botanical Garden.

---

[5] PEARSON, K., and others. Phil. Trans. Roy. Soc. London A **197**:285–379. 1901.

TABLE I

CORRELATIONS FOR FRUIT LENGTH AND FERTILITY; 1906 SERIES

Number of shrub	r_{lo}, Length and ovules per locule	r/E_r	r_{ls}, Length and seeds per locule	r/E_r	r_{lf}, Length and ovules failing per locule	r/E_r
11	−.246±.035	7.02	.376±.032	11.75	−.399±.031	12.87
12	.201±.034	5.91	.347±.031	11.18	−.093±.035	2.66
13	.300±.035	8.56	.537±.027	19.87	−.258±.036	7.17
14	.103±.039	2.63	.328±.035	9.36	−.185±.038	4.87
15	.124±.037	3.34	.247±.035	7.05	−.042±.038	1.08
16	.180±.038	4.74	.347±.034	10.19	−.083±.039	2.13
17	.198±.037	5.35	.311±.035	8.87	−.050±.038	1.30
18	.014±.039	.36	.343±.034	10.07	−.264±.036	7.30
19	.068±.039	1.75	.320±.035	9.14	−.129±.038	3.32
20	.117±.037	3.15	.533±.027	19.75	−.325±.033	9.84
21	.026±.039	.67	.439±.031	14.15	−.382±.034	11.22
22	.333±.034	9.78	.343±.034	10.10	−.166±.037	4.48
23	.096±.039	2.44	.315±.035	8.98	−.153±.039	3.91
24	.244±.037	6.58	.418±.032	13.05	−.178±.038	4.67
25	.175±.038	4.59	.475±.030	15.83	−.300±.035	8.56
26	.167±.038	4.40	.456±.031	14.70	−.225±.037	6.08
27	.247±.037	6.68	.421±.032	13.15	−.117±.038	3.67
28	.082±.038	2.14	.364±.033	11.02	−.227±.036	6.31
29	.093±.039	2.37	.415±.032	12.95	−.270±.036	7.50
30	.043±.039	1.08	.402±.033	12.18	−.297±.036	8.25

The physical constants (means, standard deviations, and coefficients of variation) for the characters dealt with need not detain us here. The correlations for the 1906 series are given in table I, and for the 1907 series in table II.[6] Assuming, as is generally done, that a correlation coefficient of 2.5 times its probable error may be regarded significant, we may examine the results for the 1906 series.

In one case r_{lo} is significantly negative; the remaining 19 constants are all positive, and 12 are greater than 2.5 E_r. There can be no doubt of the significantly positive correlation between length and number of seeds per locule. Comparing r_{lo} and r_{ls} by means of

---

[6] In calculating the probable errors for these individuals, *N* was taken as the number of locules, not as the actual number of fruits; perhaps the latter course would have been better. There is always some question as to what should be done in cases of this kind. See PEARSON, Phil. Trans. Roy. Soc. A **197**:295, footnote; HARRIS, Biometrika **7**:308–309. 1910.

TABLE II

CORRELATIONS FOR FRUIT LENGTH AND FERTILITY; 1907 SERIES

Number of shrub	r_{lo}, Length and ovules per locule	r/E_r	r_{ls}, Length and seeds per locule	r/E_r
11	.261 ± .057	4.56	.296 ± .056	5.26
17	.383 ± .046	8.41	.148 ± .052	2.84
19	−.005 ± .047	.11	.271 ± .043	6.28
24	.101 ± .038	2.64	.291 ± .035	8.24
29	.154 ± .046	3.36	.424 ± .039	10.96
31	.132 ± .038	3.53	.421 ± .031	13.42
32	.135 ± .043	3.14	.305 ± .040	7.67
33	.152 ± .039	3.88	.397 ± .034	11.75
34	.168 ± .043	3.95	.490 ± .033	14.72
35	−.022 ± .046	.48	.439 ± .037	11.94
36	.376 ± .043	8.79	.467 ± .039	11.98
37	.021 ± .040	.53	.419 ± .033	12.78
38	.040 ± .040	.99	.398 ± .034	11.84
39	.209 ± .057	3.65	.413 ± .050	8.32
40	−.255 ± .036	7.05	.334 ± .034	9.71
41	.132 ± .049	2.70	.303 ± .045	6.69

table III, we find in every case $r_{ls} > r_{lo}$, and that in 17 cases of the 20 it is significantly greater in comparison with its probable error. All the correlations for l and f are negative, and 17 out of the 20 may be safely regarded as significantly negative.

TABLE III

$r_{ls} - r_{lo}$ FOR 1906

Number of shrub	Difference $r_{ls} - r_{lo}$	$\frac{\text{Difference}}{\text{P.E. diff.}}$
11	.622 ± .047	13.23
12	.146 ± .046	3.17
13	.237 ± .044	5.38
14	.225 ± .052	4.33
15	.123 ± .051	2.41
16	.166 ± .051	3.26
17	.112 ± .051	2.20
18	.329 ± .052	6.32
19	.252 ± .052	4.84
20	.417 ± .046	9.06
21	.413 ± .050	8.25
22	.011 ± .048	.22
23	.219 ± .052	4.21
24	.174 ± .049	3.56
25	.300 ± .048	6.26
26	.288 ± .049	5.89
27	.173 ± .049	3.54
28	.282 ± .050	5.65
29	.322 ± .050	6.44
30	.360 ± .051	7.05

TABLE IV

$r_{ls} - r_{lo}$ FOR 1907

Number of shrub	Difference $r_{ls} - r_{lo}$	$\frac{\text{Difference}}{\text{P.E. diff.}}$
11	.034 ± .080	.43
17	−.235 ± .069	3.39
19	.276 ± .063	4.35
24	.190 ± .052	3.65
29	.270 ± .060	4.50
31	.289 ± .049	5.91
32	.169 ± .059	2.90
33	.245 ± .052	4.73
34	.322 ± .054	5.95
35	.461 ± .059	7.88
36	.091 ± .058	1.58
37	.398 ± .052	7.73
38	.358 ± .052	6.86
39	.204 ± .076	2.70
40	.589 ± .050	11.79
41	.171 ± .067	2.57

For the 16 plants of the 1907 series, the correlation for length and ovules is positive in 13 cases, and negative in 3; significantly negative in one instance, and significantly positive in 11. The values are low, but, as in 1906, there seems to be a slight positive relationship between the length of the pod and the number of ovules formed.

For length and number of seeds all the correlations are positive, and all are significantly positive with regard to their probable errors. Furthermore, they are of a substantial order of magnitude, ranging from 0.150 to 0.500.

Comparing, as in 1906, by taking the differences $r_{ls}-r_{lo}$, with their probable errors as shown in table IV, we find that in one case the difference is negative[7]; of the 15 positive differences, 13 are over 2.5 times their probable error and hence trustworthy.

Diagrams for r_{lo}, r_{ls}, and $r_{ls}-r_{lo}$ make the relationship very clear. Graphs for only one year (1906) need be published. In these diagrams (fig. 1) the vertical line represents the zero line on either side of which the constants would fail if they were due merely to the chance errors of random sampling. The magnitude of the constants for the individuals, or the difference in their correlations, is shown by the length, and the sign by the direction of the bars. The amount of the constant which might be due to the probable errors of random sampling (2.5 times its probable error) is shown by the unshaded area, while the shaded portion gives some idea of the biological trustworthiness of the constant. These diagrams make very clear to the eye that

1. The length of the fruit and both the number of ovules formed and the number of seeds developing are interdependent, and often moderately closely interdependent.

2. The correlation for length and number of seeds per locule is significantly higher than that for length and number of ovules per locule.

3. These two facts taken in conjunction indicate that there is some physiological relationship between the length of the fruit and the number of seeds developing.

---

[7] It is significantly negative in comparison with its probable error. I find no slip in the arithmetic. The constant is based on only 53 pods, and so too much significance must not be attached to it.

B. DISCUSSION OF DATA FROM GENERAL SAMPLES

In the foregoing section, samples from each tree were treated individually, because heterogeneity of material may have a considerable influence upon correlation. The constant for any single individual cannot be taken to typify *Staphylea trifolia* as a species. Not only is each plant in some measure individual, but the constants based upon a small series of fruits may be too large or too small by an amount known as the probable error of random sampling.

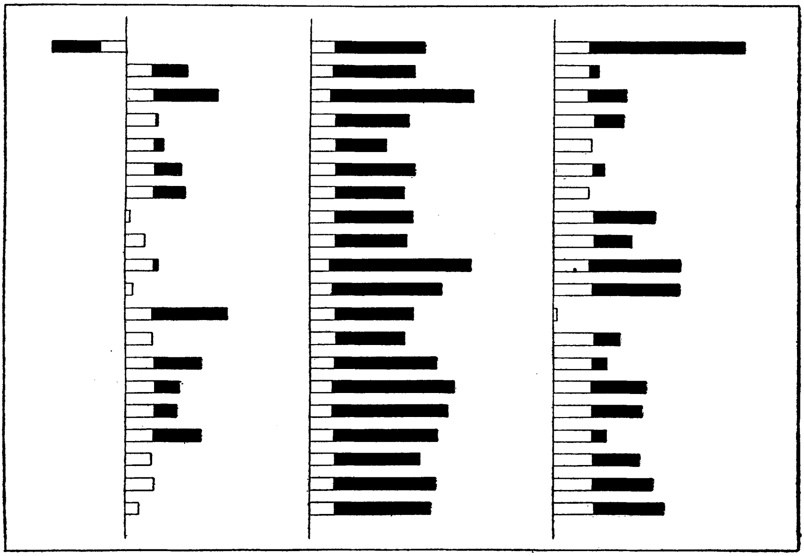

FIG. 1.—Diagram to show intensity of correlation between ovules and length (r_{lo}), seeds and length (r_{ls}), and their difference ($r_{lo}-r_{ls}$), in 20 individuals of *Staphylea;* the individuals are numbered 11 to 30 from the top of the diagram; the length of the bars shows the intensity of the correlation, and the shaded area the statistically significant amount, that is, the excess over 2.5 times the probable error.

To free our constants as far as possible from the influence of probable errors, and to gain more definite conceptions of the actual intensities of our relationships, we may (*a*) examine the means of the constants for the individuals, and (*b*) combine all our subsamples to form one large "population" of 2059 pods (6177 locules) for 1906, and 1218 pods for 1907. Table V contrasts the values obtained from the lumped samples or "population," with the (unweighted) means of the constants for the several individuals.

The means of the three characters are roughly the same for both methods.[8] The standard deviations of both l and o show large differences. The S.D. is dependent upon the differences of the means of individuals as well as upon the variation among the pods of the same plant, hence the mean S.D. of the individuals is less than that for the combined sample.

TABLE V

COMPARISON OF CONSTANTS FOR GENERAL POPULATION WITH THE MEAN VALUES CALCULATED FROM SAMPLES FROM INDIVIDUALS

	CONSTANTS FOR 1906			CONSTANTS FOR 1907		
	Population	Average	Difference	Population	Average	Difference
Mean ovules	8.1588	8.1538	.0050	7.3716	7.4532	.0816
Standard deviation of ovules	1.1244	.7673	.3571	1.2225	.8065	.4160
Mean seeds	.7149	.7165	.0016	.7427	.7418	.0009
Standard deviation of seeds	.8878	.8462	.0416	.9031	.8665	.0366
Mean length of pod	6.3356	6.3374	.0018	6.7570	6.7453	.0117
Standard deviation of length of pod	1.3663	.9773	.3890	1.3385	.9194	.4191
Correlation, length, and ovules	.0652	.1281	.0629	.0261	.1240	.0979
Correlation, length, and seeds	.3522	.3867	.0345	.2019	.3636	.1617
Correlation, length, and $\frac{\text{seed}}{\text{ovule}}$ index	.3418	.....	.....	.2018	.....	.....

Consider the correlations. For both "populations" r_{ls} is positive, and, while not large, is significant in comparison with its probable error. In both series, r_{lo} is positive, but numerically the values are too small to be of any practical significance. Both r_{lo} and r_{ls} are lower for the general sample than for the means of individuals.[9]

---

[8] Had the number of pods been the same for all plants, or the plants weighted with the number of pods taken, both methods would necessarily have given the same results.

[9] This result is at first rather puzzling. Generally the combination of several series of material raises r, but it is not necessary that this should always be the case. See footnote in a later section.

The regression straight-line formulae are:

For ovules per locule and length: for 1906, $l=5.689+.079\ o$; for 1907, $l=6.546+.029\ o$.

For seeds per locule and length: for 1906, $l=5.948+.542\ s$; for 1907, $l=6.535+.299\ s$.

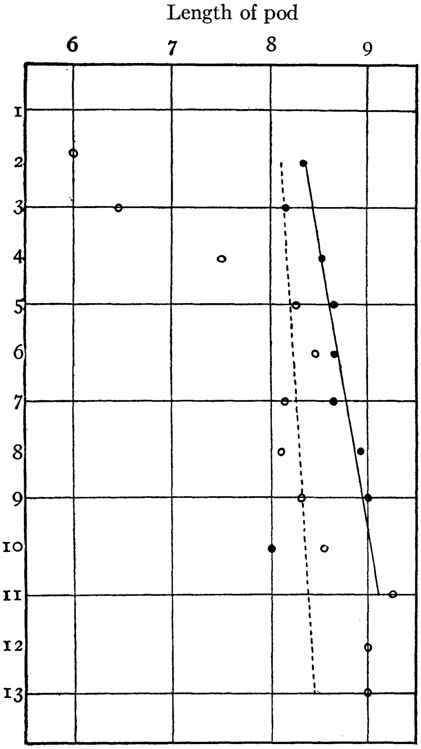

FIG. 2.—Regression of pod length on number of ovules per locule; solid dots and firm line = 1906; circles and broken line = 1907.

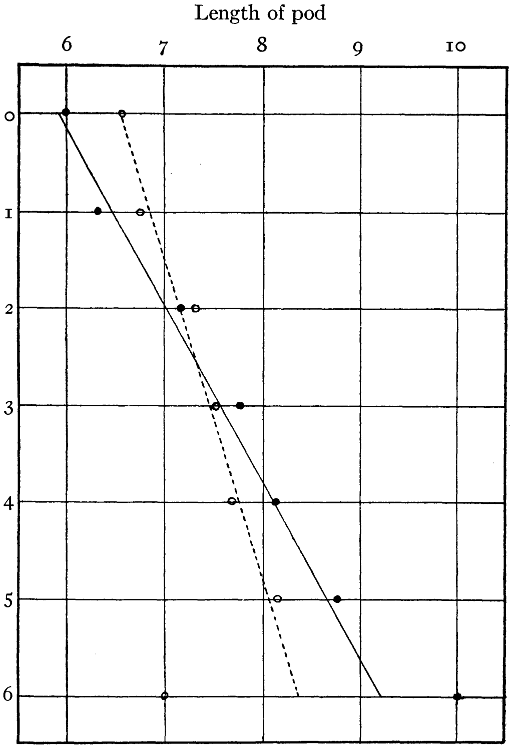

FIG. 3.—Regression of pod length on number of seeds per locule; legend as in fig. 2.

Fig. 2 shows the regression for length on ovules, and fig. 3 that for length on seeds. A straight line seems to represent fairly well the increase in pod length associated with increase in number of seeds. For ovules and length the agreement appears to be very bad, especially in 1907.[10] These statistical difficulties do not effect the fundamental conclusion to be drawn, and it is in perfect accord with that from the individual shrubs. There is a sensible correlation between length of pod and number of seeds, quite independent of the relationship between number of ovules per locule and length.

TABLE VI

Ovules formed per pod	Length of pod, 1906									
	3	4	5	6	7	8	9	10	11	Totals
15......	..	..	1	..	..	..	..	..	..	1
16......	..	..	..	..	..	..	..	..	..	...
17......	..	1	..	..	2	..	..	..	..	3
18......	..	4	11	9	2	3	..	..	..	29
19......	..	7	16	17	8	8	2	..	..	58
20......	1	8	14	24	19	12	5	..	..	83
21......	..	2	37	35	24	25	7	..	..	130
22......	..	7	30	63	33	27	7	..	..	167
23......	..	14	41	90	69	20	15	2	..	251
24......	..	35	117	167	108	57	27	7	1	519
25......	4	20	45	43	46	23	12	1	..	194
26......	..	12	26	38	34	12	9	3	..	134
27......	..	10	24	30	29	11	3	5	..	112
28......	..	6	24	24	16	15	6	5	..	96
29......	..	7	19	29	27	14	9	4	..	109
30......	1	8	26	29	36	13	5	4	..	122
31......	..	..	4	11	3	4	5	..	..	27
32......	..	..	..	..	4	2	..	1	..	7
33......	..	1	1	1	1	..	..	..	..	4
34......	..	..	1	..	1	..	..	..	..	2
35......	..	..	..	..	..	1	..	..	..	1
36......	..	..	..	..	..	1	..	..	..	1
Totals	6	142	437	610	462	248	112	32	1	2050

C. TOTAL SEEDS PER FRUIT AND FRUIT LENGTH

The normal fruit of *Staphylea* has three locules. In considering the relationship between fertility and length, should we correlate between the number of ovules or seeds maturing per locule, or should we determine the relationship between the total number per fruit and the length ? The length of the three parallel placentae determines in large measure the length of the fruit. From the purely structural point of view, it would seem logical to compute the relationships for the fertility characters of the individual locules. But in testing the assumption that the developing seed exerts some influence upon the growth of the pod, it is important to determine whether there is a closer interdependence between the length and the total seeds developing per fruit than between the length and the number developing per locule. As a basis for comparison the same relationship for total ovules per fruit must also be computed.

---

[10] One must remember that classes 7, 8, and 9 exert a powerful influence on the slope of the line, while the extreme classes are represented by so few cases that great irregularity is to be expected, even when dealing with as large numbers as we have here. While a straight line does not represent our empirical means at all well, I cannot see that an equation of a higher order would do any better. Perhaps this irregularity accounts in part for the low value of r_{lo} calculated from the general samples.

The data are: ovules and length, 1906 (table VI) and 1907 (table VII); seeds and length, 1906 (table VIII) and 1907 (table IX). The constants may at once be compared with those calculated by the other methods in table X. For total ovules and length the agreement with the same relationships for individual locules and length in the population is very close. For total seeds and length, the results are somewhat irregular.[11] The logical comparision would seem to be that of the populations rather than that of population for total seeds and the means of individuals for the numbers per locule. For both series this comparison shows a higher correlation when the total seeds developing are used.

We may now determine whether the slope of a straight line adequately describes the change in pod length associated with different degrees of fertility. Only total seeds and length need be considered. The equations are:

For 1906, $l=5.252+.505\ s$; for 1907, $l=6.186+.256\ s$.

The empirical means and the fitted lines from the above equations are shown graphically in fig. 4. It does not require a trained eye to see that the agreement is not very satisfactory. For the lower seed classes, the length of the fruit increases much more rapidly, and for the higher seed classes much less rapidly than the average rate shown by the slope of the line.

---

[11] As compared with the population constants for number of seeds per locule and length, both are considerably higher, but when compared with the means of the correlations calculated from the individuals, it appears that the constant for total seeds and length is higher than that for number of seeds per locule and length in 1906, but slightly lower in 1907.

TABLE VII

OVULES FORMED PER POD	LENGTH OF POD, 1907								
	4	5	6	7	8	9	10	11	Totals
12	..	..	1	..	..	..	..	..	1
13	..	1	..	..	1	..	..	..	2
14	..	1	3	1	..	..	..	..	5
15	1	1	2	1	..	..	..	..	5
16	..	1	5	5	4	2	..	..	17
17	3	9	13	11	9	5	..	..	50
18	3	15	28	32	24	11	2	..	115
19	..	11	24	39	22	5	2	1	104
20	2	9	23	29	26	7	3	3	102
21	..	12	29	33	18	4	3	..	99
22	2	19	40	19	19	9	3	1	112
23	5	27	33	36	8	11	3	3	126
24	3	55	91	68	31	11	5	1	265
25	2	13	39	17	12	6	2	..	91
26	1	6	14	13	2	2	..	..	38
27	1	2	5	6	7	2	..	..	23
28	..	..	1	7	5	..	..	..	13
29	..	1	2	7	6	2	1	..	19
30	1	..	2	4	6	6	3	..	22
31	..	..	..	1	3	1	..	1	6
32	..	..	..	..	1	..	..	..	1
33	..	..	..	..	..	1	..	..	1
34	..	..	..	..	..	1	..	..	1
Totals	24	183	355	329	204	86	27	10	1218

TABLE VIII

SEEDS DEVELOPING PER POD	LENGTH OF POD, 1906									
	3	4	5	6	7	8	9	10	11	Totals
0	1	5	2	2	..	..	..	..	..	10
1	5	117	333	344	127	28	2	..	..	956
2	..	17	71	175	150	59	18	2	..	492
3	..	3	19	56	103	65	24	5	..	275
4	..	..	8	19	43	41	31	7	1	150
5	..	..	3	8	26	24	12	1	..	74
6	..	..	1	1	8	16	10	5	..	41
7	..	..	..	2	2	11	8	2	..	25
8	..	..	..	1	1	3	3	5	..	13
9	..	..	..	..	1	..	2	2	..	5
10	..	..	..	2	1	..	..	3	..	6
11	..	..	..	..	..	..	..	..	..	...
12	..	..	..	..	..	1	1	..	..	2
13	..	..	..	..	..	..	..	..	..	...
14	..	..	..	..	..	..	1	..	..	1
Totals	6	142	437	610	462	248	112	32	1	2050

TABLE IX

SEEDS DEVELOPING PER POD	LENGTH OF POD, 1907								
	4	5	6	7	8	9	10	11	Totals
0.	. .	2	. .	1	. .	. .	. .	. .	3
1.	21	123	174	138	39	14	10	. .	519
2.	2	38	99	93	73	17	9	1	332
3.	. .	15	46	46	41	21	2	4	175
4.	. .	2	18	24	24	12	. .	3	83
5.	1	2	9	13	17	7	3	1	53
6.	. .	. .	5	5	4	7	1	. .	22
7.	. .	1	1	3	1	1	2	. .	9
8.	. .	. .	2	2	2	5	. .	. .	11
9.	. .	. .	1	1	1	1	. .	. .	4
10.	. .	. .	. .	. .	1	. .	. .	. .	1
11.	. .	. .	. .	. .	1	. .	. .	1	2
12.	. .	. .	. .	. .	. .	1	. .	. .	1
13.	. .	. .	. .	. .	. .	. .	. .	. .	. . .
14.	. .	. .	. .	. .	. .	. .	. .	. .	. . .
15.	. .	. .	. .	2	. .	. .	. .	. .	2
16.	. .	. .	. .	1	. .	. .	. .	. .	1
Totals.	24	183	355	329	204	86	27	10	1218

TABLE X

COMPARISON OF CORRELATIONS FOR LENGTH AND SEEDS AND OVULES PER LOCULE WITH THOSE FOR LENGTH AND TOTAL SEEDS AND OVULES PER FRUIT

RELATIONSHIP	CONSTANTS FOR TWO YEARS	
	1906	1907
For ovules and length		
Individual locules, means.	.1281	.1240
Individual locules, population.	.0652 ± .0148	.0261 ± .0193
Total ovules, population.	.0729 ± .0148	.0299 ± .0193
For seeds and length		
Individual locules, means.	.3867	.3636
Individual locules, population.	.3522 ± .0131	.2019 ± .0185
Total seeds, population.	.5888 ± .0097	.3235 ± .0173

To test this matter a little further, I have determined the correlation ratio[12] (η) for comparison with the coefficient of correlation.

For 1906, $\eta = .59724 \pm .00958$; for 1907, $\eta = .37720 \pm .01658$
For 1906, $r = .58881 \pm .00973$; for 1907, $r = .32351 \pm .01731$
$\eta - r = .00843$ $\eta - r = .05369$

---

[12] PEARSON, K., On the general theory of skew correlation and non-linear regression. Draper's Company Research Memoirs, Biometric Series, 2. London. 1905.

As constants describing the degree of interdependence between number of seeds developing and length of pod, there seems little to choose between these. To satisfy ourselves more fully concerning the rate of change in length of pod, whether it is a simple arithmetical relationship or whether it is some more complex curve, we may calculate ξ and apply BLAKEMAN'S[13] test for linearity of regression.

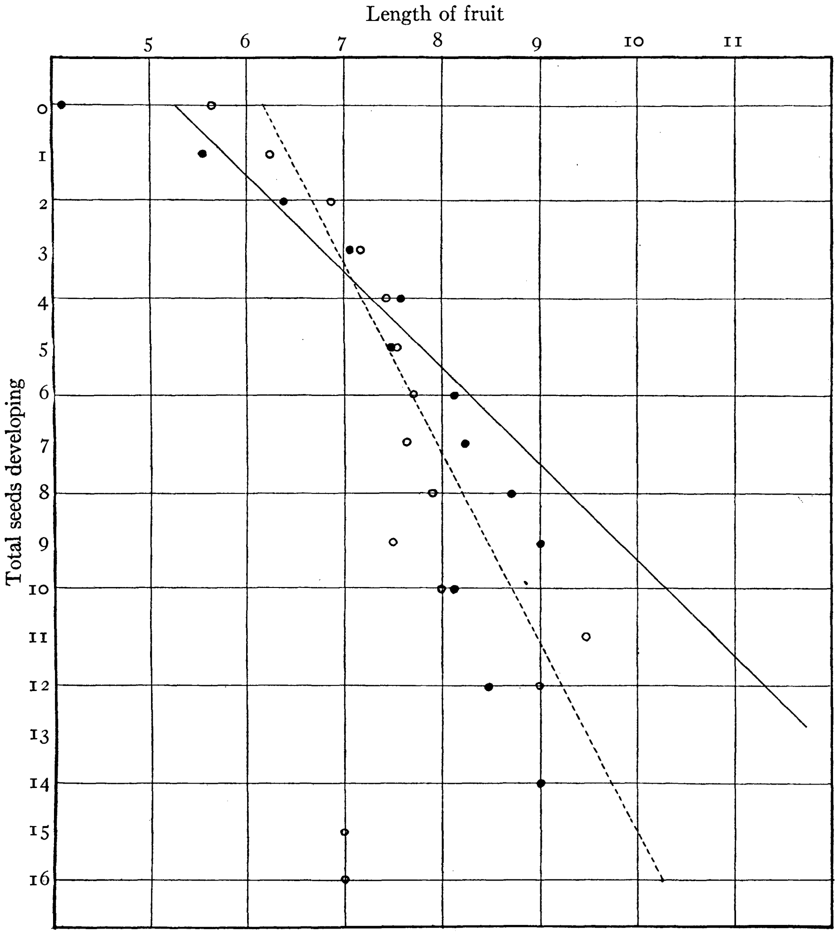

FIG. 4.—Regression of pod length on total number of seeds developing per fruit; solid dots and line = 1906; circles and broken line = 1907.

---

[13] BLAKEMAN, J., On tests for linearity of regression in frequency distributions. Biometrika 4:332–350. 1905.

$$\xi/E_\xi=\frac{\sqrt{N}}{0.67449}\cdot\tfrac{1}{2}\sqrt{\xi}\cdot\frac{1}{\sqrt{1+(1-\eta^2)^2-(1-r^2)^2}}$$

This gives

$$\text{For } 1906,\ \xi=.01000,\quad \xi/E_\xi=3.38$$
$$1907,\ \xi=.03762,\quad \xi/E_\xi=5.19$$

In both cases $\xi/E_\xi>2.5$, and regression cannot be safely regarded as linear.[14] Physiologically this indicates that after a certain number of seeds have been formed, the rate of increase in the stimulus to development exerted on the fruit falls off. Something of the same nature has been noticed in *Cercis*,[15] where it appeared that while for the central region of the seed distribution a straight line expressed the change in the pod length very well, pods with a single seed and those with the maximum number of seeds seemed a little smaller in comparison with their number of seeds than the pods of the population as a whole. When better data are available, it will be interesting to investigate this problem in greater detail, but considering the sources of error which have been emphasized above, I see no advantage in further grinding in the mathematical mill. The time would be more profitably spent in collecting other series of data for comparison. The sensibly higher coefficients for total seeds per fruit, as compared with those for seeds per locule, evidence for a direct influence of the seed upon the fruit.

---

[14] From the graphs one would almost have expected a higher value for ξ/E_ξ, but one of the difficulties in testing linearity of regression in these series is the fact that the frequencies are so concentrated into a few classes. Pods with 1, 2, 3, and 4 seeds form 91.3 per cent of the total in 1906, and 91.1 per cent in 1907. With so few pods falling in the extreme classes, it is difficult to get arrays sufficiently large to give trustworthy averages for testing the goodness or fit of means to any equation. In 1906 there are only 15 classes for 2050 pods, while in 1907 there are 17 classes for 1218 pods. Consequently the mean number of pods per array is less in 1907, and the probable error attaching to these means is greater, thus increasing the value of η by an amount depending on the magnitude of the probable errors. For 1906 the average number for arrays with entries is 157.7 pods, while for 1907 it is only 81.2. The probable errors of the means of arrays, therefore, is much higher in 1907 than in 1906, but there is no method of freeing η from their influence; perhaps these facts explain why ξ/E_ξ is greater in 1907 than for 1906.

[15] Compare the figure in BOT. GAZ. **50**:122. 1910.

CARNEGIE STATION FOR EXPERIMENTAL EVOLUTION
COLD SPRING HARBOR, N.Y.